Using artificial intelligence to assess clinicians' communication skills

BMJ 2019;364:l161 doi: 10.1136/bmj.l161 (Published 18 January 2019)

ANALYSIS

Using artificial intelligence to assess clinicians' communication skills

Most doctors have never had their communication skills formally assessed and do not know how they compare with their peers. **Glyn Elwyn and colleagues** explain how AI might facilitate this and help improve interactions with patients

Padhraig Ryan *research fellow in health informatics*[1], Saturnino Luz *chancellor's fellow*[2], Pierre Albert *PhD candidate*[2], Carl Vogel *associate professor in computational linguistics*[3], Charles Normand *Edward Kennedy chair of health policy and management*[1], Glyn Elwyn *professor*[4]

[1]Centre of Health Policy and Management School of Medicine, Trinity College Dublin, Dublin, Ireland; [2]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK; [3]School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland; [4]Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, New Hampshire, USA

Artificial intelligence (AI) has been defined as the capability of a machine to mimic intelligent human behaviour.[1] To limited extents, AI has arrived. We can give orders to our smartphones and talk to devices such as smart speakers and ask them to update us about the day's weather and traffic. They don't perform perfectly, yet the ability to understand and respond to human speech is advancing rapidly. How long might it be before speech recognition, machine learning, and other developments in AI will offer tools to medicine, and how might those tools offer insights into what happens between clinicians and patients?

Novel ways to manage practice tasks

AI research involves the development of "intelligent" computer agents. Traditionally, AI encoded existing knowledge about the world and thereby relied on prespecified human expertise. The hard coding of information into AI algorithms was typically a lengthy process. An alternative approach, known as machine learning, relies less on prior assumptions and enables computers to develop algorithms based on repeated trials and errors. Although this still requires expert human input, the time to develop AI algorithms is now often much shorter. Machine learning is prominent in tasks such as creating the equivalent of eyesight for computers, enabling self driving cars for instance.

Machine learning has potential to have a big effect on medicine,[2 3] and AI applications are beginning to emerge in healthcare. Many clinicians may prefer to use their voice rather than keyboard and mouse to interact with technology. Some, such as radiologists, already interact with digital systems using voice and physical gestures, and studies suggest important productivity gains,[4 5] although efforts are needed to ensure safety and efficiency.

When clinicians enter a diagnosis for their patient into a system, they might expect guidance about confirmatory tests and reasonable treatment options.[6-8] AI has been used to guide decisions such as the safety of combining a β blocker with a drug for arrhythmia[9] and can help clinicians diagnose late onset sepsis in premature infants.[10] Box 1 lists a few examples of AI applications.

Box 1: Examples of how AI is being developed to support healthcare delivery

- DeepMind is exploring the use of AI in managing head and neck cancer, acute kidney injury, and detection of eye disease. The ethics of using identifiable patient data from the NHS has been questioned
- Assessments are under way of the ability of the IBM Watson supercomputer to help manage cancer
- ResearchKit is a framework for mobile research using survey, biometric, and health record data
- Paige.AI is developing image processing tools to improve the accuracy and speed of pathology analysis at Memorial Sloan Kettering Cancer Center in New York

Potential role in communication with patients

An unexplored application of AI is analysis of communication in healthcare. Medical schools have all invested in teaching communication skills, but there is concern that the skills remain basic[11] and don't get much better after students qualify.[12] There is ample evidence that communications skills can be effectively developed and sustained.[13] Some specialties emphasise this during early training, notably primary care and psychiatry, using powerful methods such as simulations and feedback with video

Correspondence to: G Elwyn glynelwyn@gmail.com

recordings.[14] However, most specialties do not use these methods, and the communication skills of clinicians are often not formally assessed during training or during their years in practice.

The effect of routine assessment of clinicians' communication performance has not been studied, primarily because the analysis methods are too time consuming and expensive. However, there is consistent evidence that clinicians struggle to convey information, check understanding, and engage patients in decision making.[15 16] The lack of any effective feedback about performance makes it difficult to improve. Use of AI to assess digital recordings could provide personalised, and if necessary, confidential, detailed feedback to individual clinicians as well as comparing their performance to that of their peers. Box 2 gives some examples of the communication metrics that could be recorded and figure 1 shows how they might be presented to clinicians.

Box 2: Communication metrics that could be derived by AI analysis of consultation

Delivery

Speaker ratio—Proportion of talk by patient and clinician. Indicates willingness to listen

Overlapping talk—Interruption or simultaneous talk. Indicates respect for contribution

Pauses—Number of pauses longer than 2 seconds, which invite contribution. Indicates willingness to listen

Speed of speech—Pace can influence comprehension. Indicates wish to be understood

Energy (pitch and tone)— Influences perceptions (eg, interest and empathy)

Content

Plain language—Word choices, sentence length, and structure

Clinical jargon—Choice of terminology and effort to explain technical words

Shared decision making—Effort to inform, elicit, and integrate preferences into decisions

What might be possible?

AI could assess communication skills at much lower costs than current methods. Other sectors are already using analysis of speech and text. For example, Cogito Corp, a spin-out from Massachusetts Institute of Technology, provides call centres with voice analysis data, providing realtime feedback to employees about dominating conversations or appearing distracted. Employees receive targeted advice about how to improve their communication, and there are preliminary reports of improved performance.[17] As AI becomes better at recognising speech, irrespective of accents or language, human-machine interfaces are being mediated by talking to virtual assistants such as Siri and Alexa.

We believe speech recognition could change the way we assess clinician communication in medicine. Here we outline three areas where progress is being made by innovative groups in universities and technology companies: meaning, turn taking, and tone and style.

Meaning: analysis of words and phrases

Automated historical or realtime analysis of words and phrases could lead to innovations that were previously inconceivable. Feedback could specify whether patients and clinicians are likely to have understood each other, how aligned they were in their manner of expression, or whether the patient was encouraged to engage in shared decision making. Currently, raters assess communication elements manually using measures such as Observer OPTION-5 for shared decision making.[18] Automated analyses of recordings could accelerate these kinds of analyses and assess whether clinicians are taking appropriate histories, offering evidence based treatments, providing information to patients in accessible language that is free of jargon, eliciting patient views, and pausing to offer patients opportunities to talk. Our research team is already exploring the use of natural language processing in many of these tasks.

Eventually, just-in-time methods might analyse conversations in real time and prompt clinicians to consider diagnoses that might not be obvious or to offer a wider range of treatment options. As machine learning becomes adept at analysing realtime speech, it may become possible to assess diagnostic reasoning and the appropriateness of therapeutic decisions.

Turn-taking analysis

Analysis of turn taking could provide important insights into dialogue patterns. AI could intervene to temper kneejerk decisions to order invasive investigations—for example, cases where more detailed questioning might have led to a diagnosis of heartburn or possible muscular strain, rather than a presumption of cardiac pain.[19]

To what extent do clinicians allow patients uninterrupted time to explain their reason for the visit? What proportions of time do patients and clinicians spend speaking? Does the clinician pause to allow the patient to voice concerns or ask questions? Allowing the patient space to talk can be an act of empowerment. Turn taking can be correlated with important measures such as patients' adherence to medicines, satisfaction, and recall of information.

Tone and style in interactions

Analysis of pitch, timbre, pace, and social signals requires high quality audio recordings but could lead to many innovations and benefits to patients. We could consider questions such as whether intonation, pitch, and pace affect trust and, in turn, influence other outcomes such as patient motivation or adherence to treatment recommendations. In one research study, people's rating of the tone of voice of surgeons was associated with the likelihood of surgeons facing malpractice litigation.[20] After the implementation of appropriate data collection mechanisms, relationships such as this can be explored systematically, giving clinicians better understanding of their subtle strengths and weaknesses in communication.

In the aviation sector, pilots' key communication skills have been assessed by using algorithms to analyse their vocal pitch and energy.[21] Adaptation of such methods to the health sector might help detect high risk situations when clinicians are under stress or subjected to workloads that might affect how well they communicate. Automated analysis of pitch, timbre, and pace might support rapid detection of situations with a raised risk of potential malpractice claims. Non-verbal vocalisations may also be relevant—friendly laughter has been associated with reduced likelihood of malpractice, for example.[22]

Analysis of voice patterns could prove a rich source of information about patients' emotional states or may detect early features of illness or cognitive deficits.[23 24] Some pioneers are already exploring the detection of disease by analysing speech patterns in patients.[25] For instance, depressive episodes can be marked by systematic changes in vocal pitch,[26] and early identification of heart failure may be feasible by measuring

vocal changes arising from oedema in the vocal folds and lungs.[27]

Technical barriers

Based on current research progress, we estimate that routine, useful analysis of clinicians' intonation and turn taking will be feasible in a few years. Rudimentary content analysis, such as mapping of key clinical concepts and conversation topics, could be implemented in five to 10 years. More comprehensive content analysis, including realtime guidance on diagnosis and treatment plans, is likely to take longer.

Healthcare poses particular challenges in applying speech recognition. Clinical encounters comprise an intricate weave of at least two people speaking, often with overlapping speech. The dialogue is varied and includes greetings and partings, inquiry (history taking), explanation (for example, of diagnoses), negotiation around options, encouragement of adherence, and the provision of advice and reassurance. Skilled clinicians adjust their communication style to the needs of their patients and often use unusual technical vocabulary and similar sounding words. At present, even the most advanced AI systems are incapable of parsing and assessing the complexity of dyadic clinical interactions.

High quality digital recordings are essential for sophisticated AI analysis. This requires use of multiple microphones—for example, the clinician wearing a lapel microphone and a microphone array on a table in front of the patient. Such recordings can underpin a range of analyses, including speaker identification, semantic interpretations of words and phrases, and the assessment of vocal attributes such as pitch, pace, and timbre, which often indicate emotional states and correlate with the effectiveness of communication. Box 3 describes the progress on current technical barriers to using AI analysis.

Box 3: Technical barriers and potential solutions

Turn taking

Performance feedback requires digital recordings of clinical encounters that use more than one high quality microphone. The microphones can be embedded in a single device, known as a microphone array. The Amazon Echo device, for example, contains seven microphones. Using more than one microphone enables more accurate discrimination, because time lags between audio signals are used to separate speaker identities.[28] Details such as avoiding contact between lapel microphones and clothes would help to reduce background noise and improve sound quality. Technical solutions have been developed to automate turn-taking analysis, and research is ongoing[29]

Meaning

Turn-taking analysis is a stepping stone to more sophisticated analyses. When an AI system can identify individual speakers, the next step is to characterise the voice and words spoken; this leads to a leap in analytical potential. But first, speech recognition, where speech is converted into text by computers, must become more accurate. Advances in this area are ongoing, and some systems are near 5% error rates under specific conditions.[30 31] This is inadequate for realtime analysis of clinical care, but machines are incrementally approaching human levels of accuracy.

Even if perfect transcriptions were possible, the ability of AI systems to interpret transcribed text is limited. Computers have difficulty in interpreting linguistic ambiguity, such as identifying the referent of pronouns. Although a clinician may immediately recognise whether the pronoun "it" alludes to a medication or a diagnosis, an AI system is likely to misinterpret such statements.

Assessing non-linguistic features: pitch, timbre, pace, and social signals

Again, the analysis of vocal features relies on high quality audio files and access to large numbers of recordings to facilitate machine learning through incremental improvement of algorithms. Microphones should minimise background noise while clearly capturing the speech of interest. Non-linguistic analysis of voice is already deployed in settings such as call centres to provide realtime feedback on emotional intelligence.[32]

Possible performance feedback to clinicians

Five years ago, the idea of using AI to analyse clinical communication would have been considered impossible. However, AI developments have been rapid, driven in part by improved hardware and financial incentives to accomplish better voice and text analysis. Technology companies employ people to manually correct some errors of automated transcription devices, and this information is fed back to algorithms to improve their accuracy. There is a mass market for this technology—Apple's Siri technology, for example, handles over two billion voice commands each week.

As the field advances, research will be needed to understand whether speech analysis and personalised feedback could be an acceptable method of contributing to decision support systems or to the challenges of assessing clinician performance. For many years, researchers in healthcare communication have lamented the lack of efficient tools to assess the quality of communication skills. We may be reaching a point where speech technologies soon offer automated feedback to clinicians. It will be interesting to see whether health systems can take advantage of these latest developments and whether clinicians will be open to the use of such tools.

Key messages

- Digital recordings of healthcare interactions could lend themselves to automated analysis and machine learning methods
- Machine learning analytical methods could enable routine feedback on communication skills
- Methods for analysing aspects such as turn taking and tone are being developed
- It is unclear whether AI feedback is effective or whether professionals are willing to receive it

Contributors and sources: GE and PR initiated the article, which is part of a programme of work entitled Interaction Analytics for Automatic Assessment of Communication in Healthcare (INCA). INCA is funded by the Health Research Board, Dublin, Ireland. PR wrote the first draft, and all authors edited and approved the final manuscript. GE is the guarantor of the article.

Competing interests: We have read and understood BMJ policy on declaration of interests and have no relevant interests to declare.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Russell SJ, Norvig P, Davis E. *Artificial intelligence: a modern approach.* 3rd ed. Prentice Hall, 2010.
2 Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism* 2017;69S:S36-40. 10.1016/j.metabol.2017.01.011 28126242
3 Jha S, Topol EJ. Adapting to Artificial intelligence: radiologists and pathologists as information specialists. *JAMA* 2016;316:2353-4. 10.1001/jama.2016.17438 27898975
4 Cronin S, Doherty G. Touchless computer interfaces in hospitals: A review. *Health Informatics J* 2018 Feb 1:1460458217748342.[Epub ahead of print.]29431017
5 Hodgson T, Magrabi F, Coiera E. Efficiency and safety of speech recognition for documentation in the electronic health record. *J Am Med Inform Assoc* 2017;24:1127-33. 10.1093/jamia/ocx073 29016971
6 Middleton B, Sittig DF, Wright A. Clinical decision support: a 25 year retrospective and a 25 year vision. *Yearb Med Inform* 2016;Suppl 1:S103-16.
7 Barbieri C, Molina M, Ponce P, etal . An international observational study suggests that artificial intelligence for clinical decision support optimizes anemia management in hemodialysis patients. *Kidney Int* 2016;90:422-9. 10.1016/j.kint.2016.03.036 27262365
8 Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. *J Am Coll Cardiol* 2017;69:2657-64. 10.1016/j.jacc.2017.03.571 28545640
9 Wachter RM. *The digital doctor: hope, hype, and harm at the dawn of medicine's computer age.* McGraw-Hill Education, 2015.
10 Mani S, Ozdas A, Aliferis C, etal . Medical decision support using machine learning for early detection of late-onset neonatal sepsis. *J Am Med Inform Assoc* 2014;21:326-36. 10.1136/amiajnl-2013-001854 24043317
11 Bachmann C, Roschlaub S, Harendza S, Keim R, Scherer M. Medical students' communication skills in clinical education: results from a cohort study. *Patient Educ Couns* 2017;100:1874-81. 10.1016/j.pec.2017.05.030 28601262

12 Wouda JC, van de Wiel HBM. The communication competency of medical students, residents and consultants. *Patient Educ Couns* 2012;86:57-62. 10.1016/j.pec.2011.03.011 21501942
13 Silverman J, Kurtz SM, Draper J. *Skills for communicating with patients*. 3rd ed. Radcliffe Publishing, 2013.
14 Berkhof M, van Rijssen HJ, Schellart AJM, Anema JR, van der Beek AJ. Effective training strategies for teaching communication skills to physicians: an overview of systematic reviews. *Patient Educ Couns* 2011;84:152-62. 10.1016/j.pec.2010.06.010 20673620
15 Bernacki RE, Block SDAmerican College of Physicians High Value Care Task Force. Communication about serious illness care goals: a review and synthesis of best practices. *JAMA Intern Med* 2014;174:1994-2003. 10.1001/jamainternmed.2014.5271 25330167
16 Ahluwalia SC, Levin JR, Lorenz KA, Gordon HS. Missed opportunities for advance care planning communication during outpatient clinic visits. *J Gen Intern Med* 2012;27:445-51. 10.1007/s11606-011-1917-0 22038469
17 Matheson R. Watch your tone: Voice-analytics software help customer-service reps build better rapport with customers. *MIT News* 2016 Aug 24. http://news.mit.edu/2016/startup-cogito-voice-analytics-call-centers-ptsd-0120
18 Barr PJ, O'Malley AJ, Tsulukidze M, Gionfriddo MR, Montori V, Elwyn G. The psychometric properties of Observer OPTION(5), an observer measure of shared decision making. *Patient Educ Couns* 2015;98:970-6. 10.1016/j.pec.2015.04.010 25956069
19 Wen LS, Kosowsky JM. *When doctors don't listen : how to avoid misdiagnoses and unnecessary tests*. St. Martin's Griffin, 2014.
20 Ambady N, Laplante D, Nguyen T, Rosenthal R, Chaumeton N, Levinson W. Surgeons' tone of voice: a clue to malpractice history. *Surgery* 2002;132:5-9. 10.1067/msy.2002.124733 12110787
21 De Looze C, Vaughan B, Kelly F, Kay A, Looze D, De Looze C. *Providing objective metrics of team communication skills via interpersonal coordination mechanisms*. Interspeech, 2015.
22 Roter D. The enduring and evolving nature of the patient-physician relationship. *Patient Educ Couns* 2000;39:5-15. 10.1016/S0738-3991(99)00086-5 11013543
23 Naranjo L, Pérez CJ, Martín J, Campos-Roca Y. A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications. *Comput Methods Programs Biomed* 2017;142:147-56. 10.1016/j.cmpb.2017.02.019 28325442
24 López-de-Ipiña K, Alonso JB, Solé-Casals J, etal . On automatic diagnosis of Alzheimer's disease based on spontaneous speech analysis and emotional temperature. *Cognit Comput* 2015;7:44-5510.1007/s12559-013-9229-9
25 Place S, Blanch-Hartigan D, Rubin C, etal . Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. *J Med Internet Res* 2017;19:e75. 10.2196/jmir.6678 28302595
26 Faurholt-Jepsen M, Busk J, Frost M, etal . Voice analysis as an objective state marker in bipolar disorder. *Transl Psychiatry* 2016;6:e856. 10.1038/tp.2016.123 27434490
27 Murton OM, Hillman RE, Mehta DD, etal . Acoustic speech analysis of patients with decompensated heart failure: a pilot study. *J Acoust Soc Am* 2017;142:EL401. 10.1121/1.5007092 29092550
28 Maganti HK, Gatica-Perez D. Speaker localization for microphone array-based ASR: The effects of accuracy on overlapping speech. In: Proceedings of the 8th International Conference on Multimodal Interfaces, Banff, 2006:35-8. 10.1145/1180995.1181004
29 Church K, Church K, Zhu W, et al. Speaker diarization: a perspective on challenges and opportunities from theory to practice. In: The 42nd IEEE international conference on acoustics, speech and signal processing, New Orleans, USA. 2017. https://ieeexplore.ieee.org/document/7953098
30 Saon G, Kurata G, Sercu T, et al. English conversational telephone speech recognition by humans and machines. arXiv:170302136v1. Posted 6 Mar 2017. https://arxiv.org/abs/1703.02136
31 Hodgson T, Coiera E. Risks and benefits of speech recognition for clinical documentation: a systematic review. *J Am Med Inform Assoc* 2016;23(e1):e169-79. 10.1093/jamia/ocv152 26578226
32 Allen S. Giving voice to emotion: voice analysis technology uncovering mental states is playing a growing role in medicine, business, and law enforcement. *IEEE Pulse* 2016;7:42-6. 10.1109/MPUL.2016.2539800 27187541

Figure

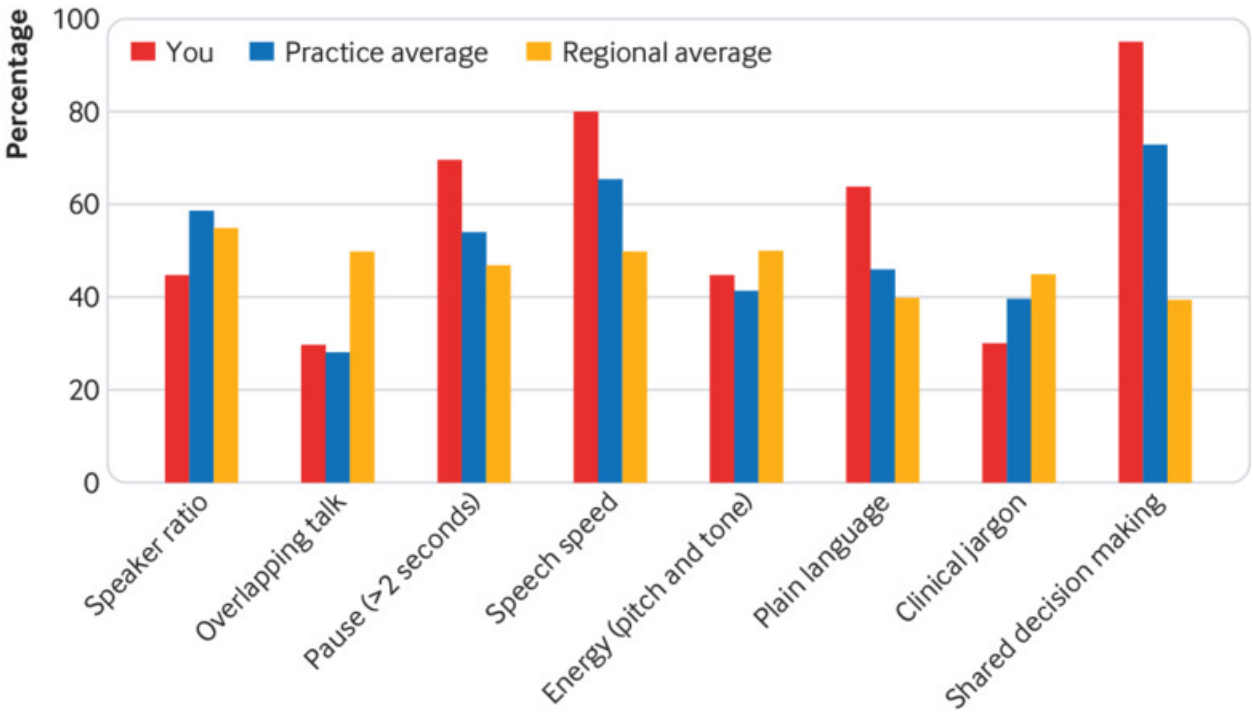

Fig 1 An imaginary communication metrics report